

## CLAIMS

- 1    1. A system configured to provide double failure-correction of two or fewer storage de-  
2    vice failures in a storage system, the system comprising:  
3        an array having a number of storage devices, wherein the number of storage de-  
4    vices is  $p$  and wherein  $p$  is a prime number greater than two, the storage devices includ-  
5    ing two storage devices selected to store redundant information, the remaining storage  
6    devices configured to store data; and  
7        a storage module of the storage system, the storage module adapted to construct  
8    the redundant information using a redundant storage algorithm involving summation or  
9    combination computation along row parity sets (rows) and diagonal parity sets (diago-  
10   nals) of the array for storage on the selected storage devices,  
11        wherein the redundant storage algorithm used to construct the two or fewer stor-  
12   age device failures is the same regardless of which storage devices fail or roles of the  
13   storage devices when constructing or reconstructing the redundant information or data.
- 1    2. The system of Claim 1 wherein the redundant information is constructed along both  
2    diagonals and rows for storage on the selected storage devices, such that neither selected  
3    storage device contains solely diagonal or row redundancy information.
- 1    3. The system of Claim 2 wherein the redundant information stored on the selected stor-  
2    age devices interacts such that the redundant information is uniquely determined by row  
3    parity and diagonal parity computation contributions.
- 1    4. The system of Claim 1 wherein the storage devices have a storage space divided into  
2    stripes and wherein any two storage devices are selected to store the redundant informa-  
3    tion, the selected storage devices varying from stripe to stripe, arbitrarily.

- 1    5. The system of Claim 4 wherein the storage module is a disk array controller.
- 1    6. The system of Claim 4 wherein the storage module is a RAID system.
- 1    7. The system of Claim 6 wherein the rows and diagonals are combined using a linear  
2    combining algorithm over bit fields that are one or more bits wide.
- 1    8. The system of Claim 7 wherein the redundant information ensures that each diagonal  
2    and each row have known parity values .
- 1    9. The system of Claim 7 wherein the roles of the storage devices in the array are irrele-  
2    vant with respect to computation of contents of any two failed storage devices.
- 1    10. The system of Claim 9 wherein the storage devices are one of video tape, magnetic  
2    tape, optical, DVD, bubble memory, magnetic disk, electronic random access memory  
3    and micro-electro mechanical storage devices.
- 1    11. The system of Claim 9 wherein the storage devices are media adapted to store infor-  
2    mation contained within data and parity blocks.
- 1    12. The system of Claim 11 wherein the array has two or more RAID groups, one or  
2    more RAID groups employing the redundant storage algorithm to protect against two or  
3    fewer failures within any one of the RAID groups, wherein each RAID group comprises  
4     $p$  storage devices with  $p$  equal to a prime greater than two and wherein a value of  $p$  may  
5    vary among the groups.
- 1    13. The system of Claim 12 wherein not all  $p$  storage devices are present in the array and  
2    wherein contents of the storage devices not present have predetermined and fixed values.

1 14. The system of Claim 1 wherein not all  $p$  storage devices are present in the array and  
2 wherein contents of the storage devices not present have predetermined and fixed values.

1 15. A system configured to provide double failure-correction of two or fewer storage de-  
2 vice failures in a storage system, the system comprising:

3 an array having a number of storage devices, wherein the number of storage de-  
4 vices is  $p$  and wherein  $p$  is a prime number greater than two, the storage devices includ-  
5 ing two storage devices selected to store redundant information, the remaining storage  
6 devices configured to store data; and

7 a storage module of the storage system, the storage module adapted to construct  
8 the redundant information using a redundant storage algorithm involving summation or  
9 combination computation along row parity sets (rows) and diagonal parity sets (diago-  
10 nals) of the array for storage on the selected storage devices, the storage module further  
11 adapted to use the redundant storage algorithm to reconstruct the two or fewer storage  
12 device failures,

13 wherein the storage devices have a storage space divided into stripes with any two  
14 storage devices selected to store the redundant information, the selected storage devices  
15 varying from stripe to stripe, arbitrarily, and wherein the two or fewer storage device  
16 failures are recovered without any knowledge of roles of those devices or any other stor-  
17 age device in the array.

1 16. The system of Claim 15 wherein the redundant information is constructed along both  
2 diagonals and rows for storage on the selected storage devices, such that neither selected  
3 storage device contains solely diagonal or row redundancy information.

- 1 17. The system of Claim 16 wherein the redundant information stored on the selected  
2 storage devices interacts such that the redundant information is uniquely determined by  
3 row parity and diagonal parity computation contributions.
- 1 18. The system of Claim 15 wherein the storage module is a disk array controller.
- 1 19. The system of Claim 15 wherein the storage module is a RAID system.
- 1 20. The system of Claim 19 wherein the rows and diagonals are combined using a linear  
2 combining algorithm over bit fields that are one or more bits wide and wherein the re-  
3 dundant information ensures that each diagonal and each row have known parity values.
- 1 21. The system of Claim 20 wherein the array is a symmetric row-diagonal (SRD) parity  
2 array.
- 1 22. The system of Claim 20 wherein the roles of the storage devices in the array are ir-  
2 relevant with respect to computation of contents of any two failed storage devices.
- 1 23. The system of Claim 22 wherein the storage devices are one of video tape, magnetic  
2 tape, optical, DVD, bubble memory, magnetic disk, electronic random access memory  
3 and micro-electro mechanical storage devices.
- 1 24. The system of Claim 22 wherein the storage devices are media adapted to store in-  
2 formation contained within data and parity blocks.
- 1 25. The system of Claim 24 wherein the array has two or more RAID groups, one or  
2 more RAID groups employing the redundant storage algorithm to protect against two or  
3 fewer failures within any one of the RAID groups, wherein each RAID group comprises

4  $p$  storage devices with  $p$  equal to a prime greater than two and wherein a value of  $p$  may  
5 vary among the groups.

1 26. The system of Claim 25 wherein not all  $p$  storage devices are present in the array and  
2 wherein contents of the storage devices not present are zero.

1 27. The system of Claim 15 wherein not all  $p$  storage devices are present in the array and  
2 wherein contents of the storage devices not present have predetermined and fixed values.

1 28. A method for protecting against two or fewer disk failures in a storage system, the  
2 method comprising the steps of:

3 providing an array with a number of disks, wherein the number of disks is  $p$  and  
4 wherein  $p$  is a prime number greater than two, the disks including two disks selected to  
5 store redundant information, the remaining disks configured to store data;

6 computing the redundant information through summation or combination com-  
7 putation along row parity sets (rows) and diagonal parity sets (diagonals) of the array;  
8 and

9 storing the computed redundant information on the selected disks.

1 29. The method of Claim 28 wherein the step of computing comprises the step of com-  
2 puting the redundant information along both diagonals and rows for storage on the se-  
3 lected storage devices, such that neither selected storage device contains solely diagonal  
4 or row redundancy information.

1 30. The method of Claim 29 wherein the step of computing further comprises the step of  
2 computing row parity along rows of the array and diagonal parity along diagonals of the  
3 array, such that the redundant information ensures that each diagonal and each row have  
4 known parity values.

1 31. The method of Claim 30 wherein the step of computing further comprises the step of  
2 computing the redundant information using exclusive-OR logical operation computa-  
3 tions.

1 32. The method of Claim 31 wherein roles of the disks in the array are irrelevant with  
2 respect to computation of contents of any two failed storage devices.

1 33. The method of Claim 28 wherein the step of computing further comprises the step of  
2 computing by addition over bit fields that are one or more bits wide.

1 34. The method of Claim 28 further comprising the steps of:  
2       dividing each disk into blocks;  
3       organizing the blocks into stripes that contain a same number of blocks in each  
4 disk, wherein each stripe comprises  $p-1$  rows of blocks; and  
5       arbitrarily assigning the roles of the blocks to thereby allow balancing of data  
6 loads across the array.

1 35. The method of Claim 34 further comprising the step of reassigning the role of a  
2 block without moving data to balance the data loads or add a disk to the array.

1 36. The method of Claim 34 further comprising the step of constructing the rows from  
2 one or more blocks from each disk, arbitrarily.

1 37. The method of Claim 34 further comprising the step of combining the rows of blocks  
2 to form a single stripe of larger blocks when  $p-1$  is a power of two.

1 38. The method of Claim 34 further comprising the steps of:  
2 adding an initialized disk to the array; and  
3 designating blocks on the initialized disk to contain data or redundant informa-  
4 tion, while ensuring that two blocks store redundant information in each stripe.

1 39. The method of Claim 38 wherein contents of the initialized disk match imputed con-  
2 tents of a missing disk, such that the redundant information in the array does not need  
3 recalculation.

1 40. The method of Claim 39 wherein the contents of the initialized disk are zero.

1 41. The method of Claim 28 further comprising the step of, wherein the disks have a disk  
2 space divided into stripes, selecting any two disks to store the redundant information, the  
3 selected disks varying from stripe to stripe, arbitrarily.

1 42. The method of Claim 28 further comprising the step of, wherein the disks are num-  
2 bered from 0 to  $m-1$ , for  $m \geq p$ , renumbering the disks by rotating the disks by amount  $k$ ,  
3 such that disk  $j$  becomes disk  $(j+k)$  modulo  $m$ , does not change the computations to com-  
4 pute the redundant information or the results of the computations.

1 43. A system configured to provide double failure-correction of two or fewer storage de-  
2 vice failures in a storage system, the system comprising:

3 an array having a number of storage devices, wherein the number of storage de-  
4 vices is  $p$  and wherein  $p$  is a prime number greater than two, the storage devices includ-  
5 ing two storage devices selected to store redundant information, the remaining storage  
6 devices configured to store data; and

7 a storage module of the storage system, the storage module adapted to construct  
8 the redundant information using a redundant storage algorithm involving summation or  
9 combination computation along row parity sets (rows) and diagonal parity sets (diago-  
10 nals) of the array for storage on the selected storage devices, the storage module further  
11 adapted to use the redundant storage algorithm to reconstruct the two or fewer storage  
12 device failures, regardless of which storage devices fail or whether the storage devices  
13 store redundant information or data,

14 wherein the storage devices have a storage space divided into stripes with any two  
15 storage devices selected to store the redundant information, the selected storage devices  
16 varying from stripe to stripe, arbitrarily, and wherein the two or fewer storage device  
17 failures are recovered without any knowledge of roles of those devices or any other stor-  
18 age device in the array.

1 44. The system of Claim 43 wherein the redundant information is constructed  
2 along both diagonals and rows for storage on the selected storage devices, such that nei-  
3 ther selected storage device contains solely diagonal or row redundancy information.

1 45. The system of Claim 44 wherein the redundant information stored on the selected  
2 storage devices interacts such that the redundant information is uniquely determined by  
3 row parity and diagonal parity computation contributions.

1 46. The system of Claim 44 wherein each storage device is divided into blocks, the  
2 blocks are organized into stripes that contain a same number of blocks in each disk.

1 47. The system of Claim 46 wherein each diagonal  $d = (i+j)$  modulo  $p$ , wherein  $i$  is a  
2 disk number of a block and  $j$  is a row number of the block within a set of  $p-1$  rows.

1 48. The system of Claim 46 wherein each diagonal  $d = (i-j)$  modulo  $p$ , wherein  $i$  is a disk  
2 number of a block and  $j$  is a row number of the block within a set of  $p-1$  rows.



- 1 49. The system of Claim 46 wherein each diagonal  $d = (i+kj)$  modulo  $p$ , wherein  $i$  is a  
2 disk number of a block,  $j$  is a row number of the block within a set of  $p-1$  rows and  $k$  is a  
3 fixed integer, and wherein  $k \neq 0$  and  $-p < k < p$ .
- 1 50. The system of Claim 46 wherein a series of blocks is organized as a chain, such that  
2 all blocks in the chain belong to two storage devices, and two blocks at either end of the  
3 chain are on different storage devices and are in the same row as nearest neighboring  
4 blocks on the chain, and wherein all other blocks in the chain are in the same row as one  
5 other block in the chain and in the same diagonal as one other block in the chain.
- 1 51. The system of Claim 43 wherein the storage module is a disk array controller.
- 1 52. The system of Claim 43 wherein the storage module is a RAID system.
- 1 53. The system of Claim 52 wherein the rows and diagonals are combined using a linear  
2 combining algorithm over bit fields that are one or more bits wide and wherein the re-  
3 dundant information ensures that each diagonal and each row have known parity values.
- 1 54. The system of Claim 53 wherein the array is a uniform and symmetric row-diagonal  
2 (SRD) parity array.
- 1 55. The system of Claim 53 wherein the roles of the storage devices in the array are ir-  
2 relevant with respect to computation of contents of any two failed storage devices.
- 1 56. The system of Claim 55 wherein the storage devices are one of video tape, magnetic  
2 tape, optical, DVD, bubble memory, magnetic disk, electronic random access memory  
3 and micro-electro mechanical storage devices.

1 57. The system of Claim 55 wherein the storage devices are media adapted to store in-  
2 formation contained within data and parity blocks.

1 58. The system of Claim 57 wherein the array has two or more RAID groups, one or  
2 more RAID groups employing the redundant storage algorithm to protect against two or  
3 fewer failures within any one of the RAID groups, wherein each RAID group comprises  
4  $p$  storage devices with  $p$  equal to a prime greater than two and wherein a value of  $p$  may  
5 vary among the groups.

1 59. The system of Claim 58 wherein not all  $p$  storage devices are present in the array and  
2 wherein contents of the storage devices not present are zero.

1 60. The system of Claim 43 wherein not all  $p$  storage devices are present in the array and  
2 wherein contents of the storage devices not present have predetermined and fixed values.

1 61. Apparatus for protecting against two or fewer disk failures in a storage system, the  
2 apparatus comprising:  
3 means for providing an array with a number of disks, wherein the number of disks  
4 is  $p$  and wherein  $p$  is a prime, the disks including two disks selected to store redundant  
5 information, the remaining disks configured to store data;  
6 means for computing the redundant information from both diagonal and row par-  
7 ity computation contributions; and  
8 means for storing the computed redundant information on the selected disks, such  
9 that neither selected disk contains solely diagonal or row parity information.

1 62. The apparatus of Claim 61 further comprising, wherein the disks have a disk space  
2 divided into stripes, means for selecting any two disks to store the redundant information,  
3 such that the selected disks vary from stripe to stripe.

- 1   63. A method for correcting double failures within data adapted for transmission over a  
2   communication medium, the method comprising the steps of:
- 3         dividing the data into packets for transmission over the communications medium;  
4         organizing the packets into one or more groups adapted to employ a uniform and  
5   symmetric double failure-correcting algorithm to protect against two or fewer failures of  
6   packets within any one of the groups, wherein each group comprises  $p$  packets with  $p$   
7   equal to a prime greater than two and wherein a value of  $p$  may vary among the groups,  
8   and wherein two packets of each group are selected to store redundant information, the  
9   remaining packets of the group configured to store data;
- 10        computing the redundant information through summation or combination com-  
11   putation along row parity sets (rows) and diagonal parity sets (diagonals) of the group;  
12   and
- 13        storing the computed redundant information in the selected packets.